Autonomous Robotic Following Using Vision Based Techniques

Robert T. Kania & Phil A. Frederick & Mike Del Rose
US ARMY RDECOM -TARDEC
Warren, MI 48397-5000

ABSTRACT

The Intelligent Systems And Autonomous Controls (ISAAC) robot is an experimental autonomous research platform being developed to advance current dismount following applications. Specifically, vision based following using pedestrian detection. The current standard in mule applications is the following of GPS waypoints. ISAAC is designed to follow a specific person using solely vision based techniques.

The core of the vision based algorithms used in this application is based on years of research from a collaboration of government and university partners. Stereo vision techniques determine a person, identify them as the leader, and map a path for autonomous following.

This paper focuses on the initialization of these algorithms and the control scheme used to implement them on a real-world platform.

Ahead of the pack

The Army is constantly looking for ways to increase the dismounted soldier's survivability, tactical awareness, and lethality. Many recent innovations have come about to do just that. AeroVironment's Micro Air Vehicle looks like a flying laptop with a propeller. It is about the same size as one, but carries a digital camera and can fly for nearly two hours. This can be used to a personal scout to keep the soldier out of harms way. Berkeley's Lower Extremity Exoskeleton straps onto a soldier's legs and lets him (or her) carry a load of 85 pounds without feeling it. This can help prevent fatigue and allow the soldier to do things they couldn't do without assistance in the past. VoxTec's Phraselator is a brick-sized one-way translation device designed for use by soldiers in countries where they don't know the language and don't have time to learn it. it uses an SD card that stores up to 30,000 common phrases useful for law enforcement, first aid, or war-fighting.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **18 APR 2006**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE
Autonomous Robotic Following Using Vision Based Techniques

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kania, Robert T; Frederick, Phil A; Del Rose, Mike

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
USA ARMY TACOM 6501 E 11 Mile Road Warren, MI 48397-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)
TACOM TARDEC

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
15693 RC

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**SAR**	**11**	

All of these examples are technologies that have obvious benefits to the soldier. This is just a small sampling of what is being developed. Show any one of these to a soldier and they'll say they want. But how many of these innovations can a soldier reasonably consider taking in the field? What equipment is the future soldier slated to carry? How can we give them all the tools they want and need and not overload them?

Tomorrow's Soldier

In 1994 the US Army launched the Land Warrior program. The Land Warrior is the vision for the soldier of tomorrow, an integrated fighting system for individual infantry soldiers which gives the soldier enhanced survivability, tactical awareness, and lethality. The Land Warrior system has many components including: a weapon system, helmet, computer, digital and voice communications, positional and navigation system, protective clothing and individual equipment.

Figure 1: Land Warrior

In February of 2005 the Land Warrior ATD (Advanced Technology Demonstration) was merged with the Future Force Warrior ATD to facilitate a more efficient spiraling of new technologies. This consolidation serves as a baseline for the Army's Ground Soldier System; a program intended to develop, test, produce, and field advanced capabilities for the future force.

The first increment equips soldiers with a dismounted battle command system that will provide situational awareness and communication known as a Commander's Digital Assistant (CDA). The CDA is fundamentally a ruggedized Compaq IPAQ linked with a SINCGARS ASIP radio for communications. It has software installed for "Blue Force" tracking similar to the functionality of the US Army Battle Command, Brigade-and-Below (FBCB2) command and control system. The system is also equipped with a joint variable message format (JVMF) database for open communications with other units, services and coalition forces.

The next increment will add the capability to interface with a Stryker. It will feature expanded situational awareness and tactical-level messaging, integrated lethality, Stryker vehicle-to-dismounted soldier communications, fratricide avoidance and battery recharging.

FCS MULE

In 2003, the Multifunction Utility/Logistics and Equipment (MULE) vehicle, developed by Lockheed Martin under contract for the FCS Lead System Integrator, The Boeing Company, was selected to move into the System Development and Demonstration (SDD) phase.

The MULE vehicle is currently designed with three variants; transport, assault and countermine. The transport configuration is designed with multiple

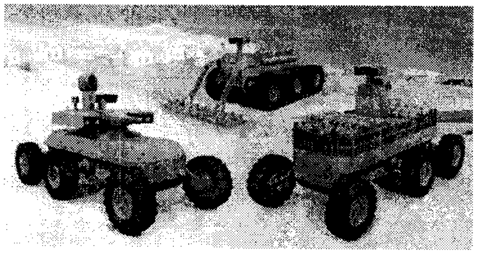


Figure 2: MULE Variants curtsey of Lockheed Martin

tie-down points and removable/foldable side railings will support virtually any payload variation. It can also serve as a med-evaq platform if needed. The assault configuration is of the Armed Robotic Vehicle – Assault (Light) (ARV-A (L)) variant. It will be armed with a line-of-sight gun and an anti-tank capability. Finally, the countermine configuration will detect and neutralize mines, as well as, mark cleared lanes.

The platform, itself, is based off of the highly mobile Spinner initiative. It has a highly advanced 6x6 independent articulated suspension with in-hub motors powering each wheel. This provides an extreme level of mobility in the most complex of terrains. The design allows the vehicle to cross gaps and climb steps of up to 1.5 meters, traverse slopes beyond 40%, and drive over obstacles up to 0.5 meters.

In the Footsteps of a Soldier

It should be apparent that the soldier of tomorrow will need the MULE vehicle to carry the plethora of equipment that they'll need/want in the field. What may not be apparent is how the soldier will interface with the vehicle. Many different control techniques have been and are being investigated – from joysticks to touch screens. Many UGVs have already been in theatre and, therefore, have been put through the ringer by soldiers in the environment they were designed (and on occasion, places the creators never imagined.)

A common theme in soldier feedback is the desire for the soldier to decide, as opposed to the machine, how much human interfacing is needed. For example, if a UGV is being used for EOD, the soldier would like to tell the vehicle to drive to the inspection site and than ignore the robot until it is in position and ready for the inspection. As of now the operator is in the loop the entire time greatly hampering their situational awareness.

In the case of MULE-type UGVs, the typical mode of control is a Leader-Follower implementation. Here, the operator traverses a path and the UGV simply follows in their footsteps at a predetermined physical and temporal distance. Thos mode of operation has the advantage of placing very little burden on the operator. The problem is that the majority of systems using this mode have a heavy dependence on GPS. A proven technology with proven limitations in the types of environments a dismounted soldier is known to operate.

Why GPS is not the long-term solution

In a theater of operations a typical war fighter can find themselves in many hostile environments. A large percentage of these environments are not very GPS friendly and many are down right intolerable. Since GPS receivers use signals from satellites they must have a clear view of the sky at all times. In proximity to tall buildings or in dense forest satellite signal may be poor or nonexistent. In order to get a three-dimensional position (latitude, longitude and height) the system

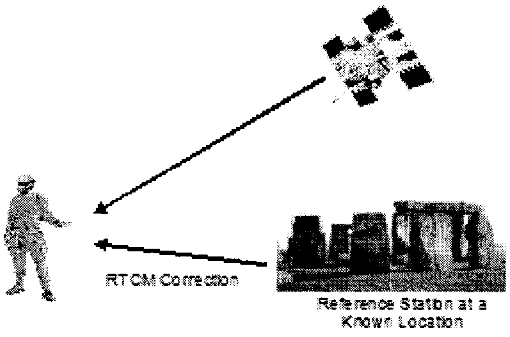


Figure 3: GPS System representation

requires a lock with a minimum of four satellites; imposing an even greater cover constraint. Also, because GPS is based on RF technology, anything that interferes with radio communications will interfere with GPS more so, because of the lower power and longer range GPS signals must travel. Finally, because most leader-follower applications require more resolution than standard GPS delivers; augmented GPS is required. This is typically thru the use of Differential GPS (DGPS) or Wide Area Augmentation System (WAAS). Both of which require an established infrastructure.

Now let's put this into historical perspective. Since the early twentieth century the United States has been at war six times: World War I, World War II, the Korean War, the Vietnam War, the First Persian Gulf War, and the Iraq War. A large percentage of the conflicts of WWI and WWII could be classified as Urban. Where as both the Korean and Vietnam Wars were fought under the canopy deep in the jungles of both countries. Neither of these environments would be considered ideal for GPS operation. This just leaves the two Gulf Wars. Therefore, two thirds of the United States most recent conflicts were in regions where GPS couldn't be relied upon. And while the open deserts of the Gulf Wars is an ideal environment for GPS usage, the required infrastructure for higher resolution GPS data is somewhat lacking.

Why Vision Based techniques are

As was previously mentioned, the current standard in mule applications is the following of GPS waypoints. This requires a considerable amount of hardware on the leader, a dependence on GPS, and an active communications link between the leader and the vehicle. None of these are tactically desirable.

However, a vision based approach could eliminate all these shortcomings. First, a fully developed vision based system would not require the leader to carry any additional hardware to interface with the vehicle. Second, this approach doesn't use GPS and therefore eliminates this dependency. Finally, because there is no additional hardware requirement, there is no need for a communications link to the vehicle. This eliminates the threat that the signal is detected and/or compromised.

These are just a few of the reasons why shifting from a GPS-based to a vision-based model for leader-follower applications would be highly desirable. By reviewing the current GPS-based model's limitations this should have become more apparent.

Meeting the Need

For over two years TARDEC has had an In-house Laboratory Independent Research (ILIR) research projects that have focused on developing technology solutions for this very problem – the *Line-of-sight dismounted leader-follower capability enhancements using pedestrian detection* ILIR. The objective of this research is to further the current developments in dismounted leader-follower technologies. Recent work done in the field of leader-follower development has not been focused on dismounted applications. This study focuses on the dismounted soldier while leveraging the technologies developed for the mounted applications – specifically research in the field of pedestrian detection. The goal is to refine pedestrian detection algorithms so that they can be used as the driving force behind dismounted following.

Vision Based Following

In 2004 we stated development of the software and hardware architecture for both a research platform and a vision system. Because this was our first venture into vision systems; we investigated various visual perception and search technologies such as: Inverse Perspective Mapping, AI Search Techniques like Breadth first, Depth First, and A* searches, Disparity Maps, Clustering, as well as Broggi's Pedestrian Detection Algorithms. For the research platform we did System/Electrical Design and Development, Mechanical Splicing, Cage Design, Finite Element Analysis on the chassis modifications, Fatigue Modeling, and Actuator Mounting.

In 2005 we focused on the problem taking pedestrian detection technology and extending it tracking and control for an arbitrary UGV. In the previous year we developed the overall algorithm to do this. This year we focused on developing the individual aspects needed to implement it. To this end we began development of a virtual environment that simulates the output of the detection technology so that we can test our vehicle control algorithms. We looked into various means of initializing the detection algorithms by researching various skin-detection and motion-detection techniques.

The complexity required of the simulation environment has grown since the initial development. We are currently able to 'walk' a virtual pedestrian in a user-defined 3D space. We are working on refining this so that we can initially simulate the detection output from Mr. Del Rose's work that will feed our control algorithms. The end-goal will be to actually give his algorithms a virtual pedestrian so that we can fully simulate the complete system that will be running on our research platform.

One skin detection technique we investigated was based on Hue and Texture extraction and processing. We used Head and Hand features to assist in height and width estimates. There were problems, however, with the Algorithm being slow. It would require tuning. Also, objects of similar HUE are susceptible to being falsely classified as skin. Finally, it does not work well in areas of high Red concentration. Another technique used training data to recognize human skin. We stored skin data in a binary image. We boxed similar groups of pixels and then found centers of each box and used a ratio to locate the pedestrian (hands & Face).

For the motion detection we worked with multiple image subtraction techniques. These worked well in the static case, but had reliability issues when we tested them on moving platforms.

The research platform, itself, is about 95% wired and the commissioning process has begun. We should be testing our vision-based following implementation with it by mid summer.

Although the intent of this work is to rely on information provided from the Human Intent and Detection (HID) Lab Pedestrian Detection System (PDS), a second PDS was investigated under this research objective. This work was accomplished in conjunction with HID Lab personnel and was initiated to provide a comparison point and potential improvements to the HID PDS system. This new endeavor explored another method to detect and track pedestrians in a set of scenes utilizing skin and motion detection. Below is a description of the two modules.

Skin Detection

This module of software was created with the intent to serve as one piece of a two module approach to detect pedestrians in a set of scenes. The objective is to infer the location of a person in a scene by segmenting it based on color and texture information. Portions of the scene with select values for these criteria are segmented out from the background and ported to a binary image.

The skin filter is based off of the Fleck and Forsyth algorithm. Utilizing a variation of the Fleck and Forsyth algorithm a RGB image is transformed to log opponent values (IRgBy) in the following manner:

$$I = \frac{[L(R)+L(B)+L(G)]}{3} \quad \text{where} \quad L(x) = 105 * \log(x+1)$$

$$Rg = L(r) - L(G)$$

$$By = (L(B) - \frac{[L(G)+L(R)]}{2}$$

Equation 1: Calculation of log opponent values from RGB values

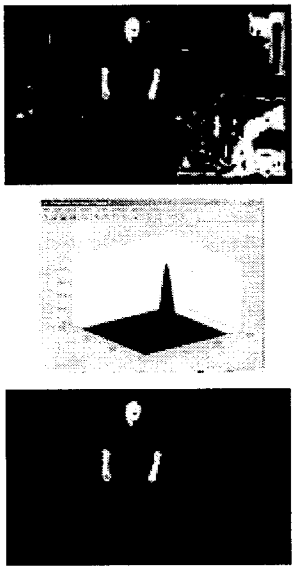


Figure 4: Skin Detection software filter results

The log opponent values are used to compute the texture amplitude, hue, and saturation values of the image. A texture amplitude map is used to identify regions of low texture information (Skin tends to have a very smooth texture). These areas of the image are identified as candidate areas for skin. The hue and saturation values are used to determine if those candidate areas of the image match the color of skin for multiple skin types in varying lighting conditions. A neural network is utilized to train the system on the hue and saturation values of these skin types in varying ambient light conditions.

Limited testing has shown the algorithm to have a positive skin detection rate of greater than 95 percent (in varying lighting conditions). However, the false positive detection rate is 12 percent. It has been discovered that the algorithm has the most trouble with false positives in areas with a high concentration of red values. Also, the training procedure to generate a hue and saturation values for various skin types in various lighting conditions has not been optimized and is a time consuming process. These two flaws must be addressed prior to implementing this system real-time on a UGV.

Motion Detection

The second software module of this PDS was the motion detection algorithm. This is a classic motion-based background subtraction technique. The assumption of this approach is that in a benign scene the background will be static while the area of interest in the foreground will be dynamic. Thus, the items of interest will change from scene to scene allowing them to be extracted from the generally static background. The first step in this procedure is to capture a picture of the static background of interest. The second step is to acquire a new image of the scene and perform image subtraction. Items that did not appear in the original static background will now be extracted due to the disparity between the two images. The third step is to update the background template with that of a new image. The frequency of the background template update should be carefully selected. Updating the background from scene to scene will increase the systems sensitivity to small changes and introduce false positives. If the background update is slow (i.e. every 10 frames) then identification of objects is not possible, as objects will move faster than the algorithm can detect them. The preliminary update rate is every 7 frames (this needs to be endurance tested). Finally, a median filter is applied to the resulting image and large areas of disparity are identified (i.e. boxed).

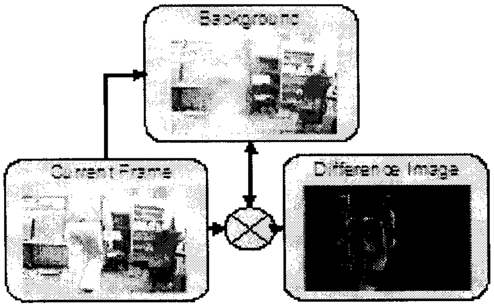


Figure 5: Motion Detection software results

This algorithm runs open loop from scene to scene (e.g. does not reference the previous image). This introduces the problem of not detecting a person when that person is standing still longer than the selected background update rate.

Combined Product (Future Work)

These two software modules can be combined to determine if an object is a person (even if that person is standing still for a long period of time). This can be accomplished by comparing the resulting binary image of each algorithm. If the objects in the resulting binary images match, with respect to time and within a given threshold, at the locations of a person's head and/or arms, then that object can be classified as human.

Also, the problem of detecting stationary people (when utilizing the motion detection algorithm) can be addressed with the combination of these two algorithms. If the previous background image from the motion detection module is stored and compared with the current resulting image from the skin detection module and current background image then it would be possible to detect standing stationary people.

Pedestrian Detection & Tracking

Pedestrian tracking was developed in house at TARDEC. The algorithm uses color blobs and disparity to dynamically update templates for tracking. Color blobs consist of grouped areas, called blobs, having similar colors. This is a convenient way to group areas together for tracking. If the frame rate is fast enough (in our experiment it is 30 frames per second) changes in color blobs from one frame to the next is gradual. This makes color blobs great for tracking a pedestrian. It even works well when the frame rate is low. Disparity is the distance from the camera (more or less by definition). It also is a convenient way of grouping similar objects based on depth.

The algorithm first reduces the image to a region of interest (ROI) that encompasses the pedestrian. Then it determines disparity and color blobs to keep for comparison to the next image. When the next frame comes in, the same procedure is done. It relies on the information of disparity of the previous frame along with location so a new ROI can be calculated. It is assumed that the final color blob information from the new frame has not changed much from the previous frame, so a comparison in the ROI of the new frame is matched up with the color blobs of the previous frame. The position with the best match is the position where the pedestrian is located in the new image.

This procedure also works well when the tracked person is occluded. As the person approaches an object that will occlude their body, it gradually changes as the person walks behind the object. Then it gradually changes back as the person exits from the occluded object. As long as the frame rate is fast enough, the changes to the blobs are gradual where the algorithm can easily pick out the new location of the person. Figure 1 shows two examples of the results from the tracking algorithm before, during, and after passing behind an object.

The key to tracking is using dynamic templates. As the new location of the person is found, the box encompassing the person (to better view the results or the operator) is now the template to be matched for the next frame, and so on... Changing the templates after each match reduces the problem to a gradual change from frame to frame. However, this does have a cost. If the object and the tracked person are within the disparity boundaries, then the object itself is considered part of the color blobs and thus used in the template to match. This problem can be reduced by limiting the boundaries to the person to only encompass him/her. More precision needs to be applied to the beginning disparity boundaries and the algorithm needs to take into account changes in distance from the cameras. This procedure should update the disparity boundaries after each frame where the changes can only be within one or two pixel differences.

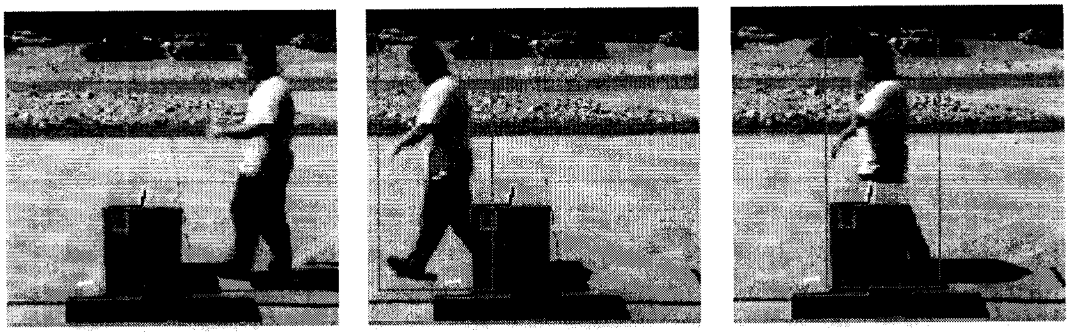
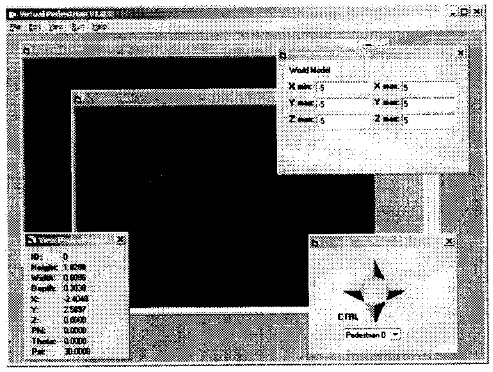


Figure 6: Output images of the tracking algorithm as the tracked person pass behind boxes

Virtual Pedestrian

The Virtual Pedestrian simulation environment is being developed to test vision-based detection and tracking algorithms in a dynamic environment. Initial work on these algorithms is usually done using recorded video. This works great for static applications. However, in the case of our leader-follower application, it is not feasible. This is because the camera does not remain at a fixed point. The camera's position will move thru free space with its position determined by the objects it's viewing. In order to test our vision-based following algorithms we had to diverge from the standard methods using recorded video.

At the core of the virtual environment is the User Defined World Model. This is basically a three-dimensional free space bounded by the user's parameters. Pedestrians and vehicles are represented as 3D boxes with dimensions set to real-world scale. These objects have headings, orientations, and velocities associated with them. There are *Field Views* where you can look into the World Model from any plane and there are *Object Views* where you can look into the World Model from the perspective of any object in the simulation.



World Model
X min: -5 X max: 5
Y min: -5 Y max: 5
Z min: -5 Z max: 5

ID: 0
Height: 1.8288
Width: 0.6096
Depth: 0.3048
X: -2.4048
Y: 2.5897
Z: 0.0000
Phi: 0.0000
Theta: 0.0000
Psi: 30.0000

Figure 7: Virtual Pedestrian Simulation Environment

The basic concept behind the simulation is simple.
Define one object as a leader and the other as a follower. Take images from the follower's forward-looking *Object View* and feed the pedestrian detection and tracking algorithms. Take the output from these algorithms to update the followers heading, orientation, and speed. If everything is working as designed you know have a follower-object that will track the leader-object in the World Model.

The current implementation is designed to work with a simplified pedestrian detection and tracking algorithm. This algorithm will simply calculate the coordinates for the 4 corners (as viewed from head-on) of the leader. These are then used by the tracker/follower algorithms to keep the follower in synch with the leader so that the 4 corners keep centered and to-scale in the follower's forward field of view.

The next step will be to wrap texture maps on the leader so that it will look more like a 3-D pedestrian and less like a box. By doing this, we hope to use the actual pedestrian detection algorithms, as apposed to the simplified versions created for the simulation.

A *Glimpse* of Things to Come!

We are finishing the simulation environment and the research platform. Upon the completion of both of these components we will be able to refine our vision-based following algorithms and test them in both a simulated and real-world environment. Once all of this is done we will be ready to build upon our technology.

One of the logical extensions of vision-based following is gesture-based control. Controlling an unmanned ground vehicle by using simple gestures would add yet another venue to the user interface community. By developing a system that uses the standard military ground guide commands any soldier could direct any such equipped UGV in the depot without the need of a specific soldier machine interface.

Another field of interest is integrating the work being developed by our *Hands-Free Tele-Operation via Physiological Signal Recognition* SBIR. This involves using a microphone in the ear canal (similar to a hearing aid) to detect sub-vocal and tongue-position commands. The current system looks like it will be easily adapted to our PDA user interface and could make for some really exciting experimentation.

Who Else Can Benefit from This?

This would help meet the objective force requirements for leader-follower capability for the dismounted soldier by enhancing and complimenting work done by the LSI. This research would not only feed into the existing Army STO and ATD efforts (Robotic Follower, Crew-integration and Automation Testbed , Human Robot Interface ,ARV Robotic Technologies, Objective Force Warrior), but also the unmanned systems (MULE and Soldier UGV) to be fielded for Land Warrior and FCS.

Although our current applications are inline with the FCS vision of the MULE and dismounted following, the technology itself is not necessarily limited to military robotics. Any scenario where you have a leader and a follower could be implemented. Consider a golf caddy that follows you from hole to hole, or a blacktop/gravel hauler for city street repair keeping step with its work force. These are just a few of the many examples that could come to mind. The implementation need not be limited to vehicle driving. This technology could be extended to assembly line work. Imagine an operator working in a virtual environment with scaled-down pieces of a product and heavy-duty multi-axis arms following their every move doing the actual assembly.

Acknowledgements

Mike "The Rose" Del Rose, US ARMY TARDEC, Pedestrian Detection and Tracking
Justin Teems, US ARMY TARDEC, Skin and Motion Detection
Matt De Minico, US ARMY TARDEC, Skin Detection

References

1. "Operator Spaceborne Path Planning for Unmanned Ground Vehicles(UGVs)", IEEE MILCON, September 2005
2. '"Autonomous Robotic Following Using Vision Based Techniques", Ground Vehicle Survivability Symposioum, April 2005
3. http://www.geocities.com/jaykapur/face.html
4. http://www.ufes.br/~marar/ARTIGO2.pdf
5. http://www.enpc.fr/certis/PapersNikos/cvpr04.pdf
6. "Insiders view to the DARPA Grand Challenge", NDIA Intelligent Vehicle Symposium, June 2004
7. "Operator Control Units for the Dismounted Soldier", NDIA Intelligent Vehicle Symposium, June 2003
8. "Multipurpose Autonomous Robotic Construct ", AUVSI's 11th Annual Intelligent Ground Vehicle Competition, June 2003.
9. "Multipurpose Autonomous Robotic Construct ", AUVSI's 10th Annual Intelligent Ground Vehicle Competition, July 2002.
10. "Robotic Follower: near-term autonomy for future combat systems", AUVSI 29th Annual Symposium and Exhibition, July 2002